Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase

Sylvie Doublié[1†], Gérard Bricogne[2,3], Chris Gilmore[4] and Charles W Carter, Jr[1*]

[1]Department of Biochemistry and Biophysics, CB 7260, University of North Carolina at Chapel Hill, Chapel Hill NC 27599-7260, USA, [2]LURE, Bâtiment 209D, 91405 Orsay Cedex, France, [3]MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK and [4]Department of Chemistry, Glasgow University, Glasgow, G12 8QQ, UK

Background: Tryptophanyl-tRNA synthetase (TrpRS) catalyzes activation of tryptophan by ATP and transfer to $tRNA^{Trp}$, ensuring translation of the genetic code for tryptophan. Interest focuses on mechanisms for specific recognition of both amino acid and tRNA substrates.

Results: Maximum-entropy methods enabled us to solve the TrpRS structure. Its three parts, a canonical dinucleotide-binding fold, a dimer interface, and a helical domain, have enough structural homology to tyrosyl-tRNA synthetase (TyrRS) that the two enzymes can be described as conformational isomers. Structure-based sequence alignment shows statistically significant genetic homology. Structural elements interacting with the activated amino acid, tryptophanyl-5′AMP, are almost exactly as seen in the TyrRS:tyrosyl-5′AMP complex. Unexpectedly, side chains that recognize indole are also highly conserved, and require reorientation of a 'specificity-determining' helix containing a conserved aspartate to assure selection of tryptophan *versus* tyrosine. The carboxy terminus, which is disordered and therefore not seen in TyrRS, forms part of the dimer interface in TrpRS.

Conclusions: For the first time, the Bayesian statistical paradigm of entropy maximization and likelihood scoring has played a critical role in an X-ray structure solution. Sequence relatedness of structurally superimposable residues throughout TrpRS and TyrRS implies that they diverged more recently than most aminoacyl-tRNA synthetases. Subtle, tertiary structure changes are crucial for specific recognition of the two different amino acids. The conformational isomerism suggests that movement of the KMSKS loop, known to occur in the TyrRS transition state for amino acid activation, may provide a basis for conformational coupling during catalysis.

Structure 15 January 1995, **3**:17–31

Key words: aminoacyl-tRNA synthetase, conformation changes, evolution, maximum entropy phasing, specificity

Introduction

The crucial role of aminoacyl-tRNA synthetases (aaRSs) [1,2] in maintaining the fidelity of the genetic code has motivated intense study of the sources of their specificity for cognate amino acids and tRNAs. The catalytic machinery for ATP-dependent amino acid activation and acyl transfer to the cognate tRNA resides in tertiary structural domains that are highly conserved within two different aaRSs classes, each consisting of the enzymes for 10 different amino acids [3,4]. Amino acid specificity evidently evolved within each class by detailed modifications of the conserved 'activation domain'. Since their efficient discrimination between competing amino acids arises within a nearly invariant tertiary structural framework, aaRS provide excellent examples of how enzyme active sites are engineered for specific binding.

Active sites in class I aaRSs, exemplified by TyrRS [5], MetRS [6] and GlnRS [7], are built around a canonical dinucleotide-binding fold first observed in dehydrogenases [8]. Associated with this characteristic tertiary structural motif are two consensus amino acid sequences, HIGH and KMSKS (using the one-letter amino-acid code) [9,10], variations of which are found in all class I enzymes. Amino acid sequences outside these so-called 'signature' sequences show considerable variability from one enzyme to another, reflecting the fact that cognate tRNA recognition requires additional, more idiosyncratic domains that vary significantly, even within the same class [11], and the fact that their most recent common ancestors date from the establishment of the genetic code. Owing to this variability, little evidence has emerged for close evolutionary relationships among synthetases for different amino acids within a class. Nevertheless, it is generally agreed that class I subclasses do exist [2,12–14]. TrpRS is thought to be more closely related to TyrRS than to any other class I enzyme, but the significance of this relationship has been questioned [13].

We show here that TrpRS strongly resembles TyrRS, as we provide the first example of two aaRSs in which tertiary structural homology extends beyond the activation domain throughout the entire structure. Sequence alignment based on structural superposition supports an unexpectedly recent evolutionary divergence. Moreover, the TrpRS active site reproduces the specific interactions with adenine, ribose and α-amino moieties previously implicated in the TyrRS catalytic mechanism [5,15]. The two enzymes also use nearly identical side-chain arrangements for amino acid recognition. This new structure therefore provides an unparalleled opportunity to examine how aaRSs achieve specificity for different amino acids.

The ligation state and crystallogenesis of any aaRSs crystal structure are potentially of considerable importance, because these enzymes appear to use conformational

*Corresponding author. †Present address: European Molecular Biology Laboratory , c/o ILL, 156X, 38042 Grenoble Cedex, France.

 ISSN 0969-2126

coupling to link the catalytic rate constant (k_{cat}) for acyl transfer to the binding of specific tRNA identity elements distant from the active sites [2,16,17]. The three previously solved class I aaRSs were each crystallized with different ligands: GlnRS with ATP and $tRNA^{Gln}$ [7], MetRS with ATP [6], and TyrRS with tyrosyl adenylate, tyrosinyl adenylate and tyrosine [5]. However, these three structures are different enough that they provide little insight into the structural basis for coupling of specificity and catalysis. Conclusions of this sort can best be drawn by examining a series of different complexes involving the same enzyme. *Bacillus stearothermophilus* TrpRS provides such a series [18,19].

We describe here the first structure in this series, that of TrpRS in complex with the activated amino acid tryptophanyl-5′AMP. This complex and its structure are unusual because they are formed enzymatically under the constraints of the crystal lattice, after the crystals are grown in complex with a different ligand. Tetragonal TrpRS crystals grow initially after making the low-molecular-weight product, tryptophanyl-2′(3′)ATP, which remains bound to the crystals [20]. Trp-2′(3′)ATP is made from the activated amino acid, Trp-5′AMP, which must be synthesized first, by acyl transfer to a second ATP molecule under the potassium phosphate crystal growth conditions. The activated amino acid ultimately observed in this structure re-forms only when crystals are subsequently transferred from that mother liquor to ammonium sulfate for data collection [19]. These crystals diffract to very high resolution (1.7 Å), and consequently 326 of the 328 residues can be positioned in the 2.9 Å electron-density map.

Results and discussion

Structure determination

The structure determination itself is significant because maximum-entropy methods [21,22] played a crucial role in solving the phase problem, which was aggravated by the loss of isomorphism upon derivatization. We succeeded by using selenomethionyl TrpRS as an isomorphous derivative, together with a new Bayesian approach [21,22] in which phase permutation is driven by a powerful combination of maximum entropy and solvent flatness [23].

TrpRS has 10 methionines per monomer. We had engineered selenomethionyl TrpRS for phase determination using multi-wavelength anomalous dispersion (MAD) but could not collect suitable MAD data without the large solid-angle coverage afforded by the use of image plates. $\{|F_{Se}|-|F_S|\}$ isomorphous-replacement data collected with an image plate using copper radiation represented a crucial source of phase information, but finding the selenium positions by difference Fourier methods required a suitable set of phases. To obtain these phases from other available heavy-atom derivatives, we had to deal simultaneously both with a serious lack of isomorphism, resulting in large starting-phase errors, and an initially poorly known molecular envelope. Because of these shortfalls, entropy maximization constrained by solvent flatness was unsuccessful as previously applied [23].

This deadlock was broken by identifying strong unphased reflections and permuting their phases according to incomplete factorial sampling designs [18]. Permutation experiments were scored using the 'log-likelihood gain', a statistic which compares the observed X-ray amplitudes with those estimated by maximum-entropy extrapolation [24]. Phase indications were then estimated jointly for all permuted reflections by least squares with t-testing. This bootstrapping procedure later also included envelope permutations, which allowed us to define the correct molecular envelope and to locate 9 of the 10 selenium atoms [25].

Combined multiple isomorphous replacement and anomalous scattering (MIRAS) phases (Table 1) produced a map with a clearly defined solvent boundary and extensive secondary structures but with poor continuity and side-chain definition (Fig. 1a). Maximum-entropy solvent flattening with additional phase permutation improved that map significantly, extending phases from 3.1 Å to 2.9 Å (Fig. 1b). Main-chain atoms throughout the molecule and virtually all side chains have well-defined density in the resulting map, which closely resembles the final, $\{2|F_{obs}|-|F_{calc}|, \phi_{calc}\}$ (Fig. 1c). TrpRS thus represents the first application of a fully fledged Bayesian phase-determination process [21] to the solution of an unknown structure, showing its feasibility and relevance. It provides a paradigm for applying maximum-entropy methods to difficult macromolecular crystal structures with medium resolution (3.0 Å) data. The usefulness of this paradigm has since been demonstrated in the difficult 3.0 Å structure determination of human chorionic gonadotropin [26].

Description of the structure

TrpRS forms an elongated α_2 dimer with dimensions 28 Å×44 Å×112 Å. Each monomer (M_r=36012 [27]) has two, well-separated domains of unequal size (Fig. 2a). A central domain largely comprises a canonical Rossmann dinucleotide-binding fold (residues 1–200). A smaller, distal lobe (residues 207–280) is formed from a four-helix bundle in which two of the three connecting loops have exceptionally high mobility.

The last 60 or so amino acids (residues 265–326) form a long, discontinuous α-helix, running from one extremity to the dimer axis (Figs 2b and 2c). It turns abruptly near the point where it emerges from the distal lobe, and folds back across the amino-terminal domain, ultimately tucking into a pocket on the other subunit. It traces a superhelical path from one end of the monomer to the other, with the symmetry-related helices in the dimer forming distinctive ridges. These two helices have no known functional significance, but are an ironic validation of the single letter code for tryptophan (Fig. 2c).

A curious network of hydrophobic interactions involving methionines 314, 318 and 322 from the carboxy-terminal

Table 1. Summary of crystallographic analysis.

	Native 1	Native 2	Derivative Semet	Derivative Au/Hg	Derivative Pb
Data collection					
Resolution (Å)	3.1	2.9	2.9	3.1	3.0
Unique reflections	8410	10092	8575	8241	8443
Total reflections	38407	44636	56731	36992	37994
Completeness (%)	99	93	80	98	91
R_{sym}(%)	4.4	6.5	7.4	7.0	8.3
Iso. differrence (%)			15	43	18
MIRAS analysis (10–3.1 Å)					
Number of sites			10	3	2
Phasing power			1.3	1.2	1.0
R_{Cullis}			0.70	0.82	0.79
Fom MIRAS	0.535				
Fom MICE	0.714				
Refinement (7–2.86 Å)					
Reflections ($I/\sigma > 2.0$)	8859				
Total reflections	9631				
R-factor (all reflections; %)	19.7				
R_{free} (%)	27.6				
Number of atoms	2617				
Rms bond length (Å)	0.011				
Rms bond angle (°)	2.8				

Abbreviations: MIRAS, multiple isomorphous replacement and anomalous scattering; MICE, maximum entropy in a crystallographic environment. $R_{sym} = \Sigma|I - <I>|/\Sigma I$, where I = observed intensity and <I> = average intensity obtained from multiple observations of symmetry-related reflections. Mean fractional isomorphous difference (iso. difference) = $\Sigma||F_{PH}| - |F_P||/|F_P|$ where $|F_P|$ is the protein structure-factor amplitude and $|F_{PH}|$ is the heavy-atom derivative structure-factor amplitude. Phasing power = rms($|F_H|/E$) where $|F_H|$ is the heavy-atom structure-factor amplitude, and E is the residual lack of closure. $R_{Cullis} = \Sigma||F_{H(obs)}| - |F_{H(calc)}||/\Sigma|F_{H(obs)}|$ where $|F_{H(obs)}|$ is the observed heavy-atom structure-factor amplitude, $|F_{H(calc)}|$ is the calculated heavy-atom structure-factor amplitude, and the summation includes only centrosymmetric reflections. Fom, figure of merit. R-factor = $\Sigma_{refls}||F_{(obs)}| - |F_{(calc)}||/\Sigma_{refls}|F_{(obs)}|$, where $|F_{(obs)}|$ is the observed structure-factor amplitude and $|F_{(calc)}|$ is the calculated structure-factor amplitude.

helix and Met92 from the same subunit at the interface helps to hold this helix against the three helices that form the bottom of the tryptophan binding site (near residues 80, 120, and 160 in Fig. 2a). None of these methionines is present in the *Escherichia coli* TrpRS [28]. The cold-sensitive behavior of selenomethionyl-TrpRS crystals suggests that this novel interaction, which we call a 'methionine zipper', may help to stabilize the thermophilic TrpRS tertiary structure. Selenomethionyl-TrpRS crystals rapidly lose all diffraction at temperatures below 12°C, whereas natural TrpRS crystals have an extended lifetime in the X-ray beam at 4°C. Selenomethionine is more hydrophobic than methionine, and cold-induced destabilization of the methionine zipper would be expected as a result of the decreasing contribution of the hydrophobic effect at lower temperatures [29]. This may account for the abrupt loss of diffraction when selenomethionyl-TrpRS crystals are cooled.

Superimposing the four class I enzyme Rossmann folds [30] leaves root mean square (rms) deviations between corresponding C_α atoms of 1.69 Å (TyrRS/TrpRS),

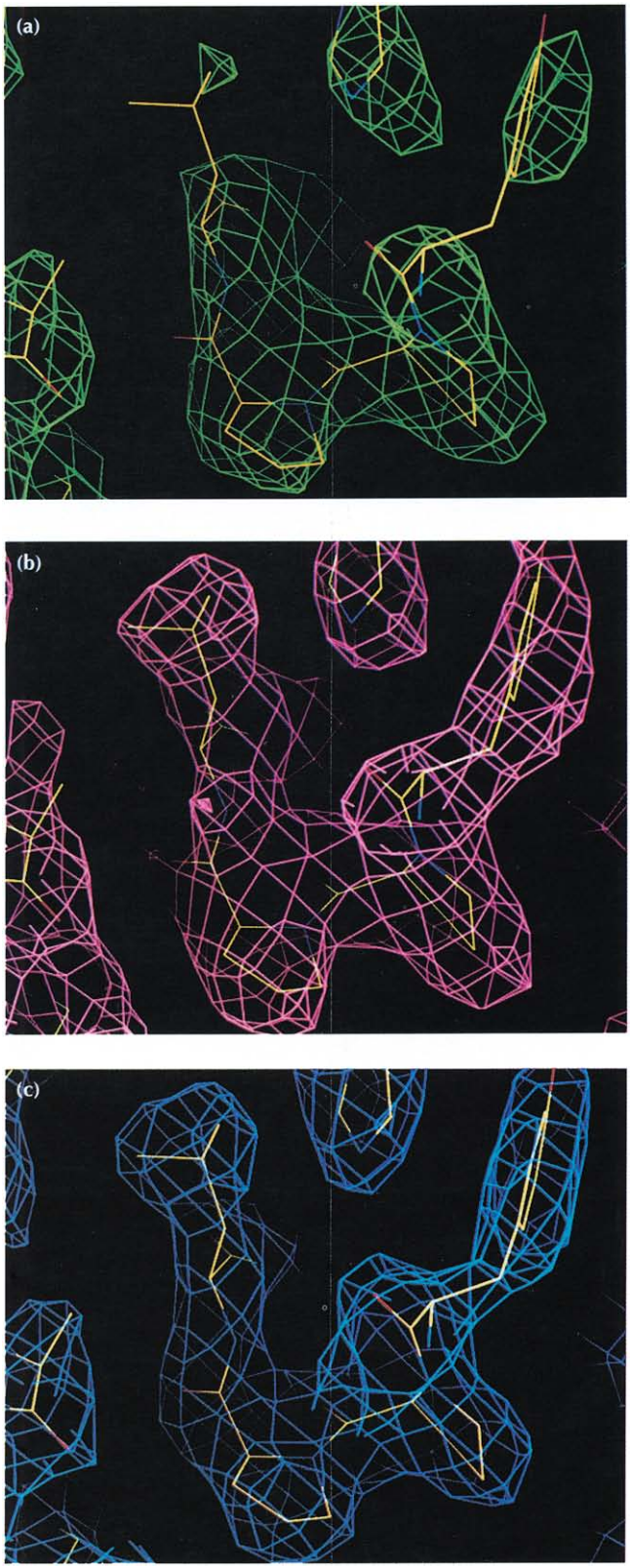

Fig. 1. Electron-density maps from the phase determination. **(a)** The best MIRAS map, phased with isomorphous differences for semet TrpRS, Au/Hg, and Pb derivatives. **(b)** The centroid map, calculated with Sim weights and phases from maximum-entropy solvent flattening and phase permutation. **(c)** The final, refined, $\{2|F_{obs}|-|F_{calc}|, \phi^{ME}\}$ map. Maps (b) and (c) are nearly indistinguishable, whereas map (a) displays serious errors, including main-chain discontinuities.

Fig. 2. (a) Stereoview of TrpRS down the dimer axis. Every 20th residue is colored red, and sequence numbers are shown for accessible multiples of 20. The two monomers are colored blue and gray. Division of the monomer into two domains is evident from the coiled cleft about two-thirds of the way out from the dimer interface. The two central cross-over connections of the Rossmann dinucleotide-binding fold are shown in darker shades. Sequences of particular importance are shown in lighter colors; these include the two signature sequences, HIGH and KMSKS (cream, between residue labels 20 and 200) and loops with exceptional mobility ($B_{C\alpha}$>40 Å^2; 115–119, 223–227 and 261–263; yellow). The tryptophanyl moiety of Trp-5′AMP is coloured green, the AMP moiety white. **(b)** Space-filling representation of the domains in the dimer, viewed from 180° about the vertical axis in (a), showing the carboxy-terminal helical segments (yellow and white) running from the distal domains to the inner edge of the nucleotide-binding fold and making contacts across the dimer axis. **(c)** View with the dimer axis vertical. The carboxy-terminal helical segments are shown in darker shades; signature sequences and the loops with high mobility are highlighted as in (a).

1.99 Å (GlnRS/TrpRS) and 2.20 Å (MetRS/TrpRS), revealing considerable similarity between the activation domains of all four class I structures. Notably, however, neither TrpRS nor TyrRS has the left-handed crossover connection found in MetRS and GlnRS [6]. The interactions between both signature sequences (^{15}TIGN^{18} and ^{192}KMSKS^{196}, cream-colored and near residue labels 20 and 200 in Fig. 2c) and the AMP moiety are discussed below.

Bound ligand and the active site for activation

TrpRS has the apparently unique ability to aminoacylate ATP [31], forming tryptophanyl-2′(3′)ATP (Trp-ATP). This unusual product is presumably formed by acyl transfer to a second molecule of ATP bound in the site normally occupied by the 3′ adenosine of tRNA^{Trp}. Radiolabeling of crystals with [^{14}C]tryptophan and [^{32}P]ATP demonstrated binding of Trp-ATP to the crystals [20], and we expected to use its location to mark the 3′-CCA tRNA-binding site, and provide evidence as to which of the two ribose hydroxyl groups is aminoacylated by *B. stearothermophilus* TrpRS [3].

An $\{|F_{obs}|-|F_{calc}|,\phi_{calc}\}$ difference Fourier map derived using only the protein atoms in the calculated model (Fig. 3a) revealed a bi-lobal region of density from which we could deduce the location of the active site. The connection between the two lobes can only be interpreted as the 5′ linkage of the adenylate intermediate, Trp-5′AMP. The radiolabeled crystals that contained the product Trp-ATP were grown and stabilized in potassium phosphate buffer; those used here were grown in the same way but subsequently stabilized in ammonium sulfate, which causes a phase transition that changes the unit cell parameters significantly (from 62.3 Å, 62.3 Å and 220.5 Å to 60.7 Å, 60.7 Å, 233.9 Å [19]) and improves the diffraction limits from 3.0 Å to 1.7 Å. Breakdown of Trp-ATP apparently occurs in ammonium sulfate, and is followed by re-synthesis of the intermediate, leading to a significant repacking of lattice contacts.

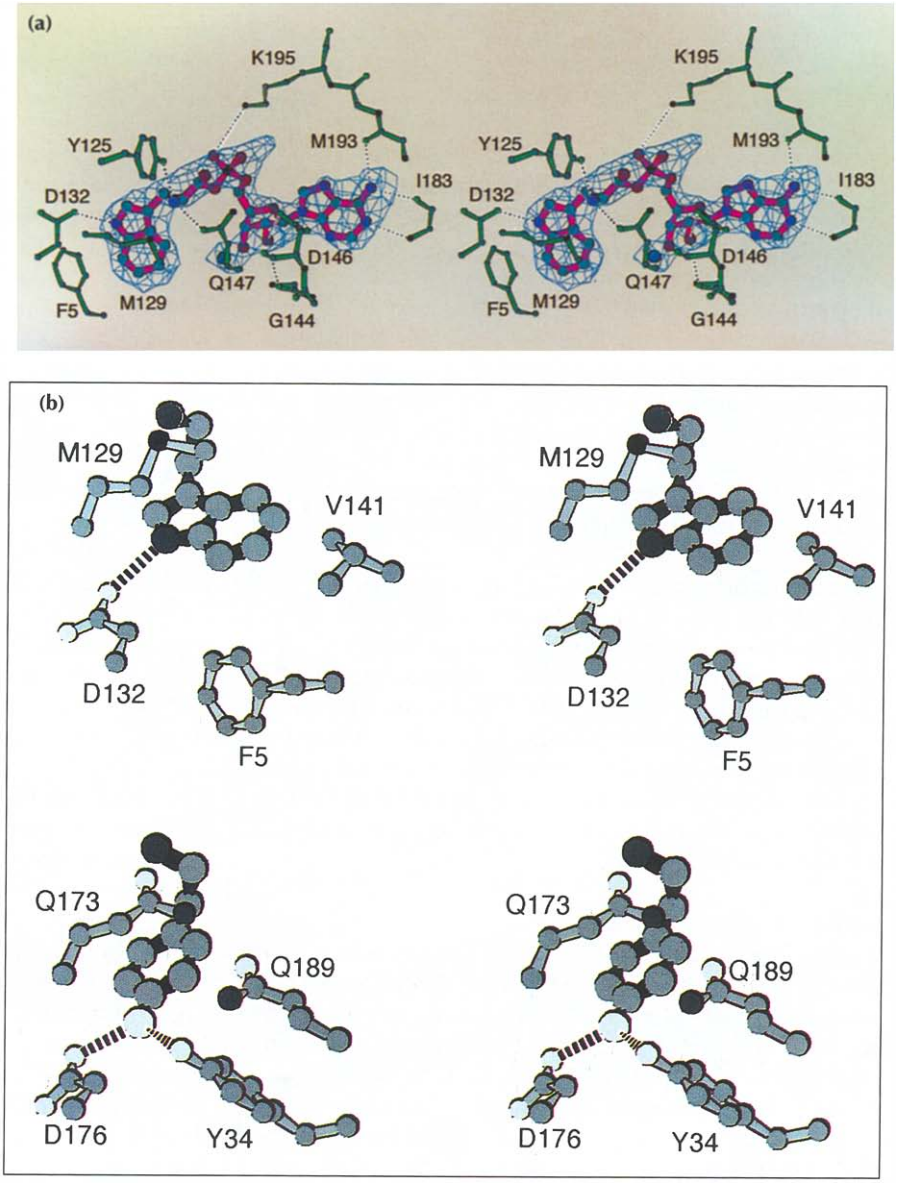

Fig. 3. **(a)** Stereoview of the $\{|F_{obs}|-|F_{calc}|,\phi_{calc}\}$ difference map with Trp-5′AMP, contoured at 2σ. Specific interactions between TrpRS and the activated amino acid are indicated. A strongly bound water molecule found near the 3′-hydroxyl group of the ribose in both TrpRS and TyrRS complexes is shown as a blue sphere. **(b)** Stereoviews of specific interactions of TrpRS (top) and TyrRS (bottom), with the indole and phenol rings of their substrates, tryptophan and tyrosine.

Trp-5′AMP binds in the expected manner [5,7] to the Rossmann fold. One lobe of the difference density rests on Gly17, the other lies close to the second crossover connection, in the place occupied by the tyrosine moiety in the TyrRS·tyrosyl-5′AMP structure. Gly17 occurs in the sequence TIGN, which is the equivalent of the HIGH signature sequence. The glycine in this sequence is strictly conserved in class I aaRSs; it serves as a platform for binding of the adenine ring in the GlnRS·ATP·$tRNA^{Gln}$ [7,17] and TyrRS·tyrosyl-5′AMP [5] structures, as it does here.

Interactions are made with both five-membered and six-membered rings of indole in the tryptophanyl adenylate (Fig. 3b). The thioether sulfur atom of Met129 fits into the center of the five-membered ring, making van der Waals contact with each atom, while the carboxylate of Asp132 forms a charged hydrogen bond to the indole nitrogen atom. The six-membered ring stacks against the edge of the Phe5 side chain as described by Burley and Petsko [32].

Several independent lines of evidence suggest the importance of these interactions for specificity. First, selenomethionyl TrpRS is approximately twice as active as the native enzyme (S Doublié and CW Carter Jr, unpublished data). Substituting selenium for sulfur in Met129 could affect its interaction with the five-membered ring, accounting for this increased specific activity. Second, studies with fluorescent tryptophan analogs show that the adenylate of 7-aza-tryptophan is highly fluorescent, whereas that of 5-hydroxy-tryptophan is quenched, relative to its fluorescence in solution [33]. The orientation of the indole in the binding pocket is consistent with these observations; C7 interacts with the phenyl side chain of Phe5, blocking access to solvent hydrogen-bonding groups, and C5 is close to the Ser6 hydroxyl group, a potential quencher via hydrogen bonding to the 5-hydroxyl group of 5-hydroxy-tryptophan.

Interactions of substrate tryptophan with Met129 and Asp132 bear a striking resemblance to functionally significant clusters of the same three amino acids in two other proteins. Trp191 in cytochrome *c* peroxidase (CCP) interacts with a methionine (Met230) via the five-membered ring, and its indole nitrogen is hydrogen bonded to a buried aspartate (Asp235). Mutagenesis has shown the methionine and carboxylate to be essential for Trp191 to function as a free-radical intermediate in CCP-catalyzed electron transfer [34].

MetRS may use a similar triad, including conserved residues Trp305 and Glu241, to validate the selection of substrate methionine [35]. The triad differs somewhat, however; Met129 is replaced by glutamine in yeast mitochondrial and beef TrpRSs [36]. Nevertheless, the reciprocal use of different members of this triad to verify the involvement of the remaining members in correct substrate binding suggests that the interactions involving the triad are chemically significant, and that reciprocity among similar side-chain/substrate combinations may occur in aaRSs for other pairs of amino acids.

TrpRS and TyrRS are approximately conformational isomers

TyrRS has a 30-residue amino-terminal extension, missing in TrpRS, that wraps around the nucleotide-binding fold, and the carboxy-terminal 100 residues in the TyrRS structure are disordered, so their fold cannot be compared with the carboxy-terminal helix in TrpRS. Elsewhere, however, TrpRS and TyrRS share nearly complete tertiary structural homology. The interactions (detailed below) comprise a nearly one-to-one correspondence in both sequence and function, which verifies the importance of previous work with TyrRS [5,15]. This uncanny similarity prompted us to investigate just how closely the two sites can be compared by superposition. The residues of interest, including the adenylate itself and residues interacting with it, were excluded from the superposition, as were non-homologous loop sections. In particular, we excluded an entire helix, residues 122–133, which we call the 'specificity-determining helix'. The rms deviation of the remaining 57 C_α atoms from the Rossmann fold (residues 2–11/31–40, 16–19/46–49, 22–25/51–54, 38–43/68–73, 53–66/93–106, 75–81/118–124, 94–99/140–145, and 143–148/191–196, where the first pair of residue numbers corresponds to TrpRS and the second to TyrRS) was only 1.1 Å, underscoring the structural similarity and facilitating a rather precise comparison of the amino acid side-chain binding sites, which were not used in fitting.

Having superimposed the active-site domains, it became apparent that the remaining domains in TrpRS also closely resemble corresponding domains in TyrRS, although they are oriented rather differently. Two additional structural regions can be distinguished in both enzymes: the dimer interfaces, and the distal helical domains. Their structures are strikingly similar (Fig. 4) and superimpose quite well. For residues at the dimer interface 105–113/151–159 the rms deviation is 1.2 Å. For residues 209–220/267–275, 240–247/252–260, 247–258/277–285 and 262–269/293–301 in the distal helical domains the rms deviation is 2.6 Å. These superpositions give rise to extensive structure-based amino acid sequence alignments encompassing 272 residues (Fig. 5).

The distal helical domains and dimer interfaces assume different orientations with respect to the activation domains, so that the dimers appear to have rather different structures (Fig. 6). The orientation at the center of Fig. 4 was defined by superimposing the 57 residues from the Rossmann folds in a monomer, and was used as the common orientation of the lower (gray) monomers in Fig. 6. From this orientation the two carboxy-terminal helical domains within the same monomer differ by a rotation of ~17° around an axis through the amino-acid-binding site which closes the TrpRS structure, relative to TyrRS. This rotation is accompanied by collapse of the loop containing the KMSKS signature sequence, which

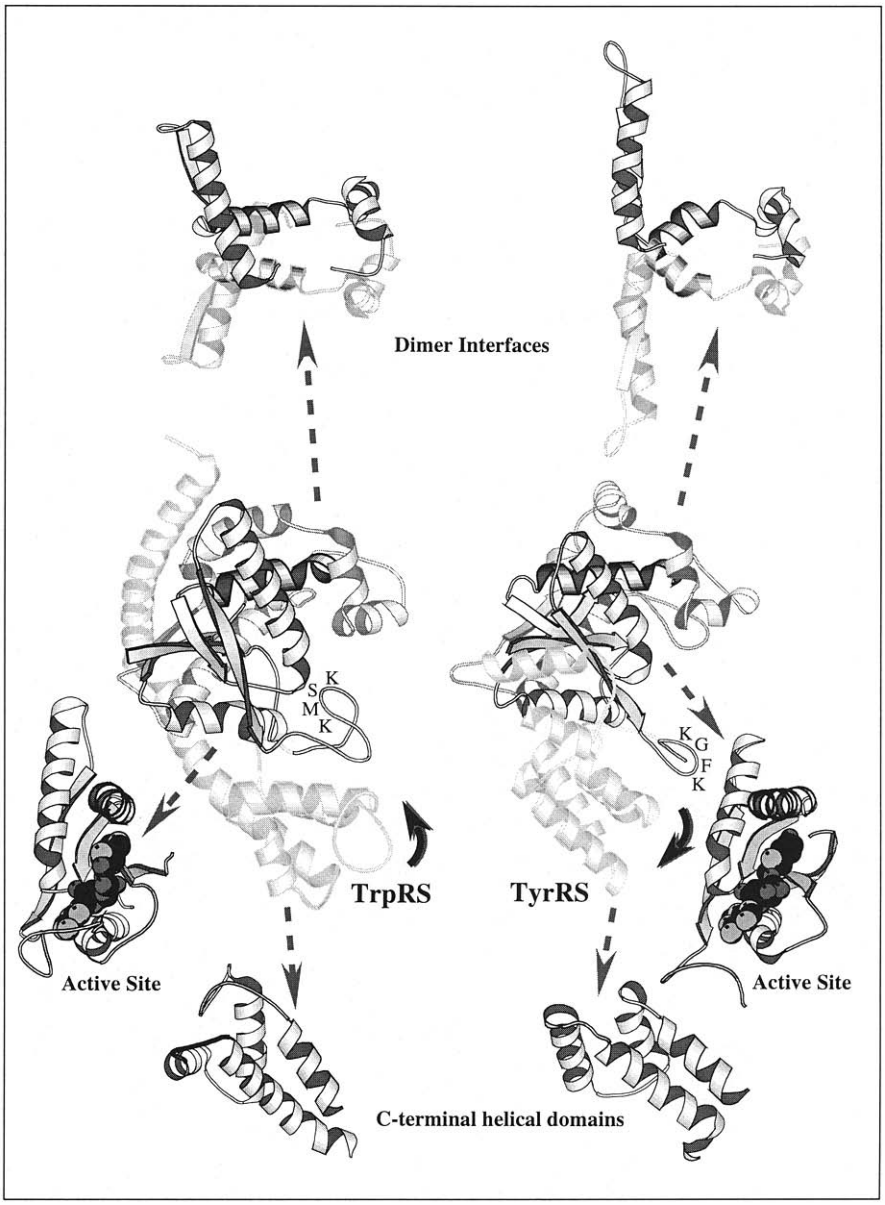

Fig. 4. Comparison of TrpRS (left) and TyrRS (right) monomers and their components. Views of the monomers (center) emphasize their similarly oriented Rossman dinucleotide-binding folds (dark lines and shading). The dimer interface (upper right background) is drawn in medium gray and the helical domain (bottom) in light gray. A 30-residue amino-terminal segment in TyrRS (mid-left foreground) and the C-terminal helix in TrpRS (left background) are also drawn in light gray. Signature sequences KMSK (TrpRS), and KFGK (TyrRS) are marked to indicate loops whose movement is implicated in pyrophosphate binding by TyrRS in the transition state [15]. Curved arrows represent 'closed' (TrpRS) and 'open' (TyrRS) configurations of this loop and the nearby helical domain in the two enzymes. Fragments drawn in pairs around the periphery have arbitrary orientations chosen, after superposition, to emphasize their similarity. Left and right: active site domains including the two crossover connections of the Rossman fold and the activated amino acid (space-filling spheres). The KMSKS sequences (not labeled) are on the lower left. Top: dimer interfaces with the molecular dyads oriented horizontally. Fragments from the second monomer are shown in gray. Bottom: carboxy-terminal helical domains.

is 'open' in TyrRS, onto the active site. Superimposing the dimer interface segments in the second (blue) monomer from the reference orientation requires a counterclockwise 17° rotation about an axis roughly parallel to the long axis of the molecules, and an additional 13° rotation about the molecular dyad axis. The effect of the combined rotations is to bring the active-site cavity opening in TrpRS closer to the orientation of the corresponding cavity in the lower monomer. The active sites of the second monomers consequently point in quite different directions in the two enzymes.

The major difference between TrpRS and the major, ordered fragment observed for TyrRS therefore consists of the relative rotations of structural components with very similar structures. In this sense they can be described surprisingly accurately as conformational isomers, despite the fact that they are different enzymes. No other pair of known class I enzymes shows such extensive structural homology. The fact that this 'isomerism' also involves a radical change in the position of the KMSKS loop, known from mutagenesis studies to move during catalysis [37], suggests that this description may be functionally relevant, as discussed below.

Interactions with the activated amino acids Trp-5′AMP and Tyr-5′AMP

The interactions with the C_α substituents, ribose, phosphate and adenine of tryptophanyl-5′AMP in the TrpRS complex are nearly identical to those which were observed for the tyrosyl-5′AMP in the TyrRS complex [5] and which have been strongly implicated in catalysis by directed mutagenesis studies [15]. Almost invariably, interactions observed in TyrRS are present in the TrpRS complex and *vice versa* (Fig. 7). Conserved interactions include: Gln147 and Tyr125 with the α NH2 group (Gln173 and Tyr169 in TyrRS); the Gln9 backbone amide with the phosphate (Asp38 in TyrRS); Asp146,

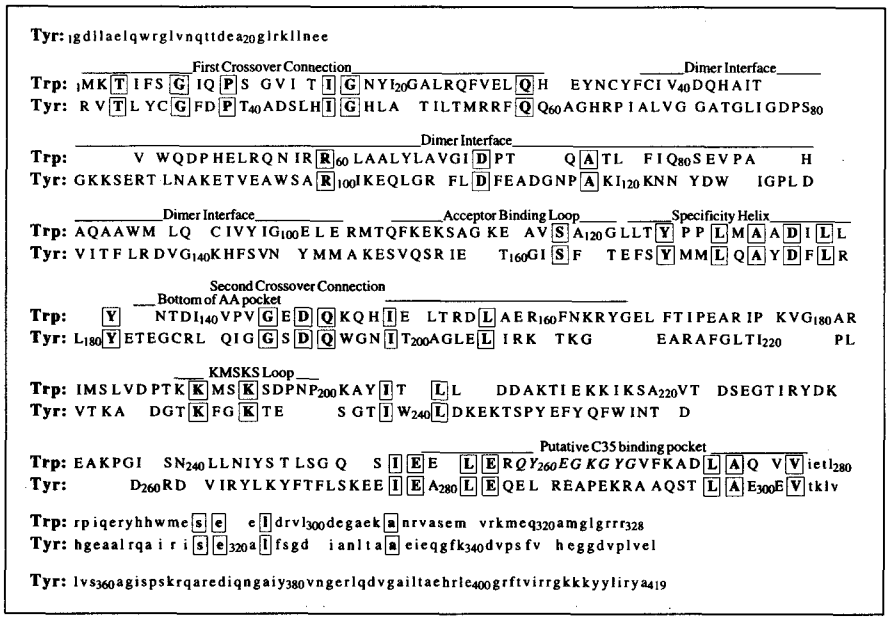

Fig. 5. Structure-based alignment of TrpRS and TyrRS amino acid sequences. Identities are in bold type. The putative binding site for C35 in the $tRNA^{Trp}$ anticodon is shown in italics. The approximate physical locations of sequences are identified above the sequences, for reference with the structural illustrations and Fig. 7.

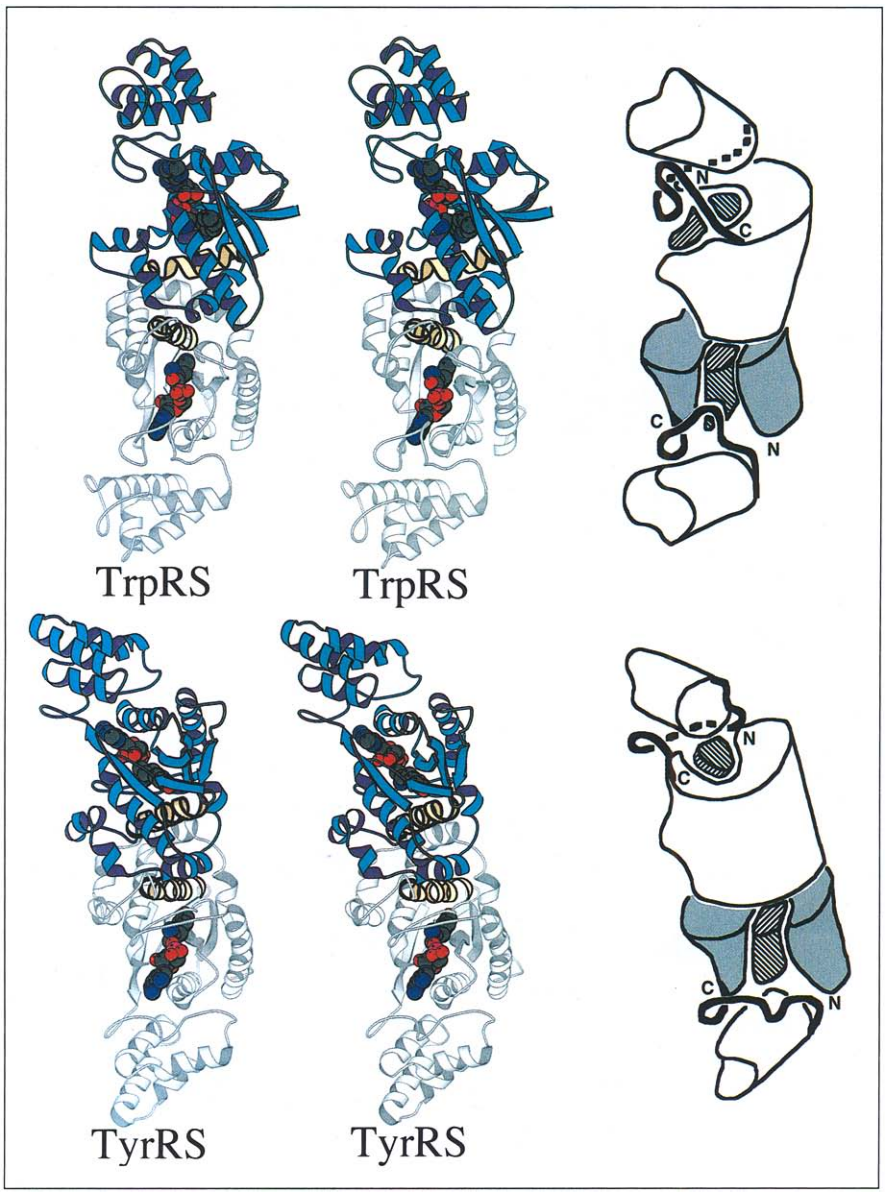

Fig. 6. Comparison of structurally homologous domains of the TrpRS and TyrRS dimers. The bottom monomer in each case (gray) is oriented such that the active sites are exactly superimposed, as in Fig. 4, giving an rms deviation of 1.1 Å. Both the dimer interface and the helical domains of the top monomers (blue) are rotated substantially, relative to each other, such that the active sites differ by ~25° and the helical domains by ~45°, as indicated in the schematic diagram to the right. The specificity-determining helices are emphasized by bold lines and by their cream color to provide a reference for the intersubunit rotation at the dimer axis. The amino-terminal and carboxy-terminal crossover connections of the Rossmann dinucleotide-binding fold flank the (activated) aminoacyl-adenylates (hatched), as indicated for both subunits in the schematic diagram. The dimer axes of both figures are nearly horizontal and in the plane of the figure. The position of the KMSKS loop is indicated for both subunits by the bold lines connecting the active site and helical domains. The helical domains and KMSKS loops are both 'closed' in TrpRS (rotated inwards and towards the dimer axis) and 'open' in TyrRS (rotated outwards and away from the dimer axis).

a bound water and the Gly144 backbone amide with the 2′-hydroxyl and 3′-hydroxyl groups of the ribose (Asp194, water and Gly192 in TyrRS) and Asn18 (His48 in TyrRS) from the HIGH sequence with the ribose.

Equally interesting are the differences between the two active sites. TyrRS has a nine-residue insertion including Asp78, which makes an additional interaction to the α NH2 group. Met193 and Ile183 in TrpRS interact with the adenine 6-amino group via their backbone carbonyl oxygens. These interactions, which include the first to be identified involving the methionine from the KMSKS consensus sequence, would preclude using GTP for activation. Because the two segments containing these differences move substantially in generating the TyrRS transition state [15], the correlation between the differences evident in Fig. 7 and the conformational isomerism implies that the two structures may represent different stages of catalysis, despite having corresponding ligands.

That possibility is reinforced by an interaction in TrpRS between the invariant Lys195, of the K(192)MSKS(196) loop and the adenylate α-phosphate, as was also seen in the GlnRS·tRNA^{Gln}·ATP complex [38]. Movement of this loop by ~7.5 Å (the average value for the four residues, KMSK) in the TrpRS complex folds it over onto the active site so that Lys195 can interact directly with the α-phosphate of the adenylate (Figs 3a, 6 and 7).

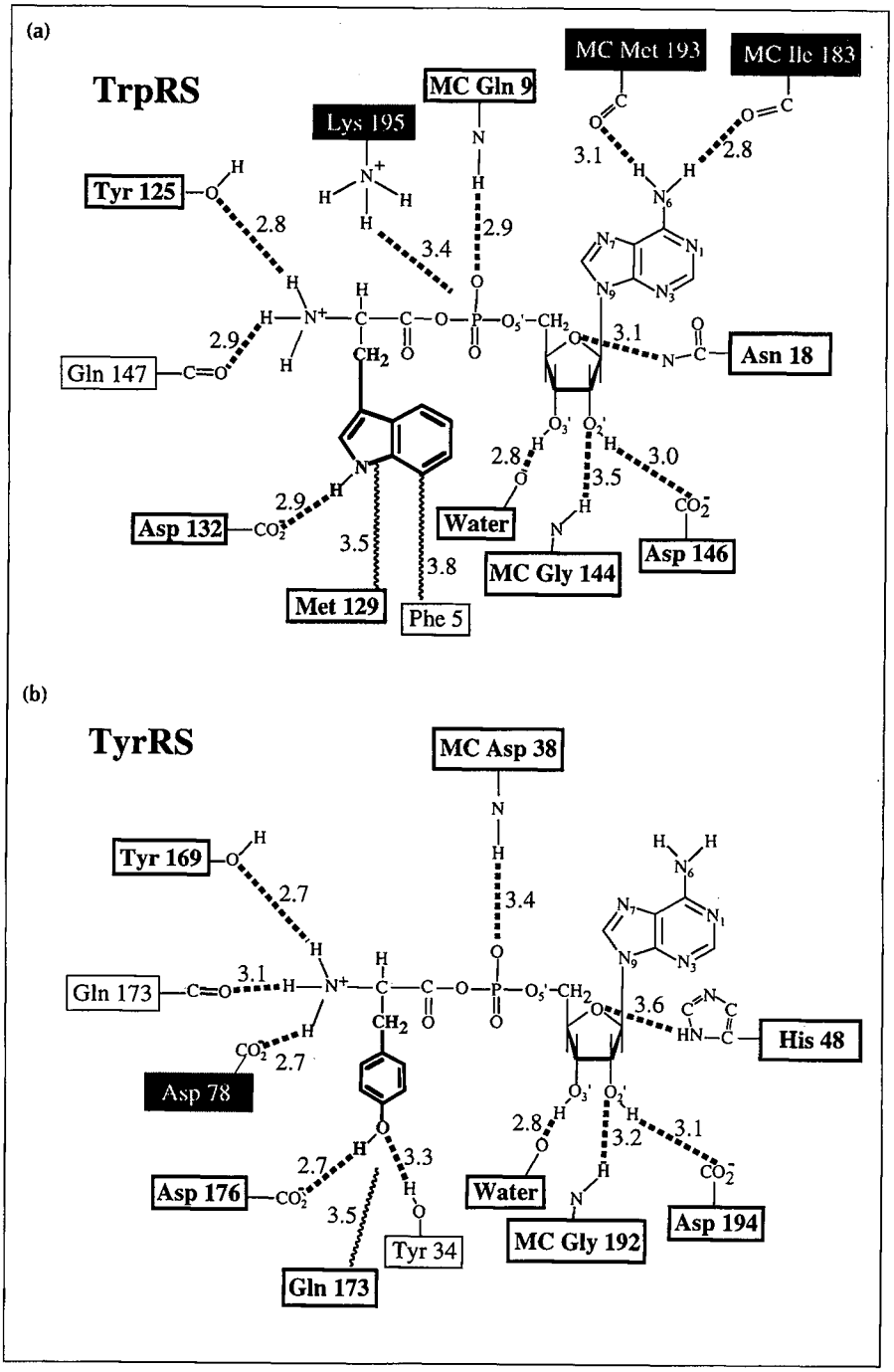

Fig. 7. Active site–ligand interactions in **(a)** the TrpRS·Trp-5′AMP complex and **(b)** the TyrRS·Tyr-5′AMP complex (after Fersht *et al.* [37]). Dashed lines with distances (in Å) indicate hydrogen bonds identified by Brick *et al.* [5], by site-directed mutagenesis [15], and from our crystal structure. Hydrophobic interactions are shown by wavy lines. Bold type denotes amino acids that are structurally superimposible and located in homologous TrpRS and TyrRS sequences. A shaded background indicates non-superimposable residues.

This is one of two loops identified by Fersht *et al.* [37] as undergoing an induced-fit motion, in particular to bring the lysine residues into range to bind the pyrophosphate moiety during the transition state for amino acid activation. In TyrRS the loop is ~8 Å too far away for either lysine residue to interact with the leaving pyrophosphate group. It is substantially closer to the putative site of the leaving pyrophosphate group in TrpRS. The other mobile loop identified by Fersht *et al.* involves Lys82 and Arg86, and is truncated in TrpRS (Fig. 5).

Substrate specificity

The biologically relevant differences between the two active-site structures concern amino acid recognition. Here, the two structures also show unexpected structural similarity (Figs 7, 8 and 9). Indeed, the sequence identity between structurally aligned segments is far more significant in the region devoted to amino acid recognition than anywhere else in the sequence (Fig. 8). Residues Tyr125, Met129 and Asp132 from the specificity-determining helix in TrpRS correspond exactly to residues Tyr169, Gln173 and Asp176 from the corresponding helix in TyrRS, which are also highly conserved in several other class I enzymes [13]. The side chains superimpose almost exactly, except for Asp132/Asp176, as described in more detail below.

Specific side chains responsible for amino acid recognition are nearly identical for the two different amino acids, and they occupy nearly superimposable positions! Whence then does the ability to discriminate originate? Part of the answer appears to be that subtle tertiary structural changes, together with differences in the depth and orientation of the binding pocket, are sufficient to preclude activation of the incorrect amino acid, despite the overall similarity of the chemical features (aromatic ring and hydrogen-bonding group) recognized by the two enzymes.

A conserved aspartate (132 in TrpRS and 176 in TyrRS [5,15]; Fig. 9a) provides a key specificity determinant in both cases because of the hydrogen bond it forms with the substrate. Recognition of the correct partner is accompanied by hydrogen-bond lengths of ~2.8 Å. In contrast, binding of tryptophan by TyrRS would lead to an inappropriate separation of 4 Å, whereas binding of tyrosine to TrpRS would involve a close van der Waals contact of ~1 Å. The likely negative charge on the carboxylate at physiological pH increases the discrimination for formation of the correct hydrogen bond [39]. A subtle difference in the shape of the active site in TrpRS arises from a 7° re-orientation of the specificity-determining helix axis, relative to the rest of the nucleotide-binding fold (Fig. 9b). Two proline residues, Pro126 and Pro127, near the amino terminus of the helix may induce a kink, leading to this re-orientation. Its effect is to bring Asp132 closer to the substrate tryptophan, in order to make the hydrogen bond to its indole nitrogen.

Side-chain packing in the interior of the pocket is specifically tailored to the respective substrates (Fig. 9b).

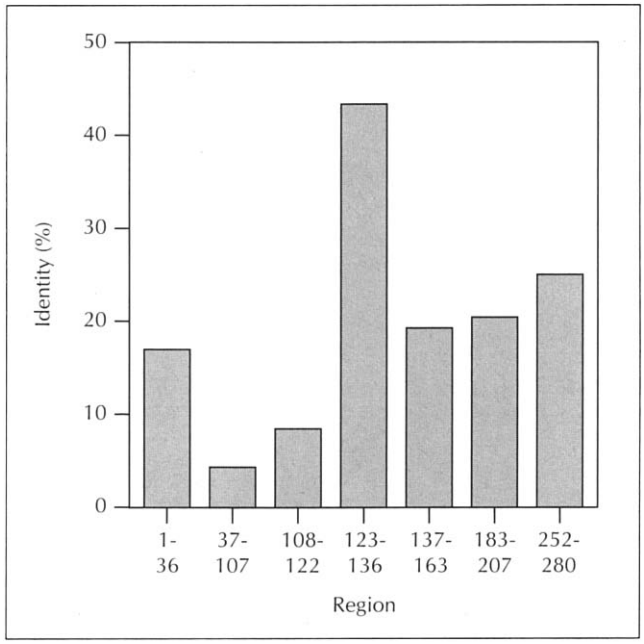

Fig. 8. Relative amino acid sequence conservation between TrpRS and TyrRS. The regions are defined as follows: residues 1–36, first crossover; 37–107, dimer interface; 108–122, acceptor-binding loop; 123–136, specificity-determining helix; 137–163, second crossover; 183–207, KMSKS loop; 252–280, helical domain. Surprisingly, the regions devoted to side-chain binding specificity (residues 123–136) are the most conserved.

Gln189 in TyrRS is replaced by the smaller and more hydrophobic Val141 in TrpRS, providing extra room only in TrpRS for the indole six-membered ring. The aromatic side chain near the amino terminus (Phe5 in TrpRS, Tyr34 in TyrRS) is also oriented differently, and in TyrRS the phenolic hydroxyl group donates a hydrogen bond to the substrate tyrosine.

Similarity of the active sites further suggests specific mutations (Phe5→Tyr, Val141→Gln and replacement of Pro126 and Pro127) that might effect a significant conversion of the TrpRS amino-acid specificity in favor of tyrosine. Such experiments are in progress (D Söll, personal communication).

Constraints on tRNA binding

One of the most interesting questions raised by any aaRS structure concerns how it recognizes its cognate tRNA. We can formulate preliminary, but significant, hypotheses about the tRNA-binding mechanism from the TrpRS dimensions and the observations detailed below.

The intact carboxy-terminal domain of TrpRS has been positioned and refined, thus the anticodon and acceptor stem binding sites cannot involve invisible or disordered segments, and should be present in our structure, unless some part of the ordered TrpRS structure undergoes a substantial rearrangement to form the tRNA-binding sites. In contrast, roughly 100 residues are missing from the TyrRS structure [5].

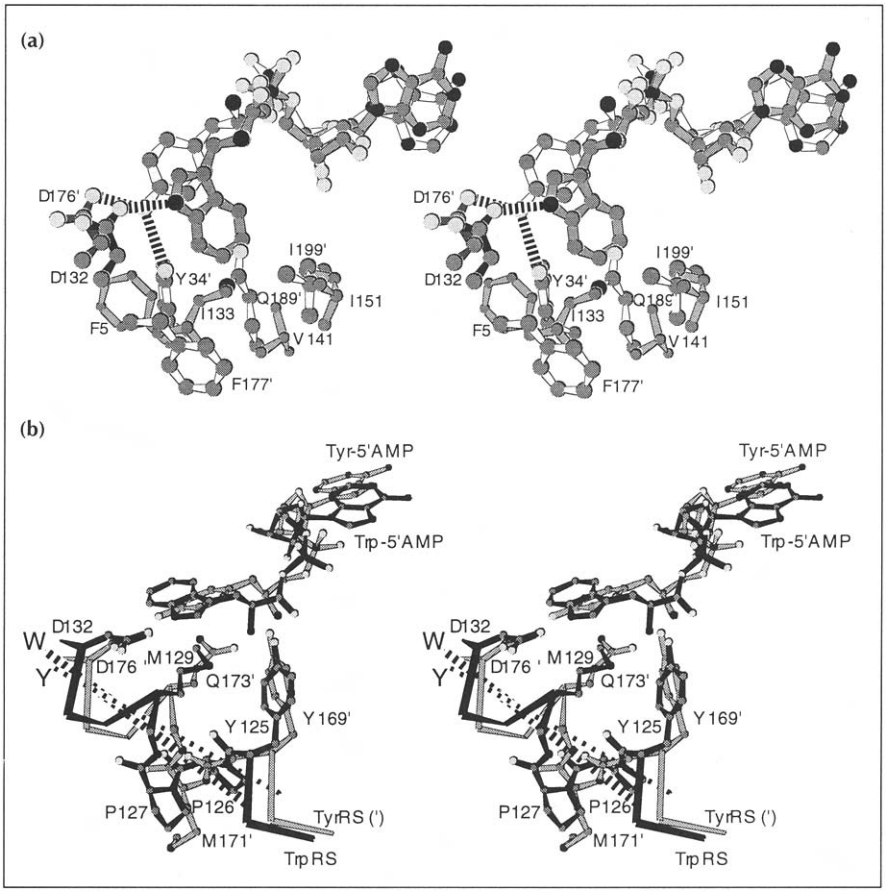

Fig. 9. Superposition of TrpRS (dark shading) and TyrRS (light shading, primed residue numbers) active sites. Hydrogen bonds between the conserved carboxylate groups of Asp132 (TrpRS) and Asp176′ (TyrRS) are shown in **(a)**. Note also the nearly exact superposition of the two ribose groups. **(b)** Re-orientation of the 'specificity-determining helix' that provides nearly all specific interactions with the respective substrates. The β-sheet of the Rossmann fold (not shown) lies above the amino-acyl-5′AMP. Dashed lines represent the axes of the specificity-determining helices ; they differ by ~7°, leading to a slight, but critical repositioning of the aspartate carboxylate groups shown in (a).

Anticodon bases C34, C35 and A36, together with G73 are major identity elements, whereas base pairs A1–U72, G5–C68 and base A9 are minor identity elements for prokaryotic TrpRS in *E. coli* [40,41] and *B. subtilis* [42]. TrpRS must then bind the $tRNA^{Trp}$ acceptor stem near the acyl group of Trp-5′AMP while the anticodon binds elsewhere. The maximum distance from the acyl group in one active site to the far edge of the same monomer is ~40 Å. Anticodons and acceptor stems are located 50–70 Å apart in the known tRNA crystal structures, thus precluding binding of all identity determinants to the same monomer.

A connecting polypeptide of ~100 residues inserted between the two halves of the Rossmann fold forms the acceptor-binding domain in GlnRS [7], and is thought to have the same function in MetRS [11]. Corresponding fragments in TyrRS and TrpRS are involved in the dimer interface. In TrpRS, the region corresponding to the GlnRS 'acceptor-binding domain' consists only of the loop 113–119 adjacent to the active site. It is disordered (B-factor ~40–60 $Å^2$; Fig. 2a) and may therefore bind to the 3′-acceptor terminus of $tRNA^{Trp}$.

Significant sequence identity exists between residues 517–524 that recognize anticodon base U35 in GlnRS and residues 259–266 in the TrpRS distal lobe:

B. stearothermophilus	TrpRS	^{259}**Gln**Tyr**Glu**GlyLys**GlyTyr**Gly^{266}
E. coli	GlnRS	^{517}**Gln**Phe**Glu**ArgGlu**GlyTyr**Phe^{524}

This loop and one facing it in TrpRS (residues 223–227; Figs 2a and 2c) have high B-factors, indicating flexibility and making them likely candidates for the $tRNA^{Trp}$ anticodon-binding site. Despite the fact that these sequences are embedded in very different secondary structures in GlnRS and TrpRS, a reasonable argument can be presented that they play similar roles in the two proteins. The corresponding base in the tryptophan anticodon is a cytidine, C35, which is distinguishable from uridine only by the amino group C4 substituent, which is a carbonyl in uridine. The only residue from the GlnRS sequence that interacts with the U35 base is Gln517; it forms a hydrogen bond to N3, which is common to cytidine and uridine. Two other residues, crucial to the GlnRS interaction, have homologs in the TrpRS structure. Arg520 interacts with the 3′ phosphate of U35. Arg341 interacts with the O4, and is a specificity determinant for the carbonyl oxygen group. In the TrpRS structure the Oε1 oxygen of Gln259 is flanked in a similar manner by Nζ of Lys263 and Oε1 of Glu255. The latter group corresponds to Nη2 of Arg341 in GlnRS, consistent with the change in specificity of the C4 pyrimidine-ring substituent from a carbonyl oxygen in U35 (GlnRS) to an amino group in C35 (TrpRS). The three key atoms from the two proteins superimpose with an rms deviation of 1.7 Å, which is well within the range that can be accommodated by side-chain rearrangements. The TyrRS structure that is homologous to residues 259–266 in TrpRS has the apparently unrelated sequence, ELREAPEKR, and there is no evidence for a homologous sequence elsewhere in the TyrRS sequence.

Both TrpRS and TyrRS lack a 'left-handed crossover connection' [6,11]. This motif, which helps position $tRNA^{Gln}$ [7] on GlnRS, is also found in MetRS, and can be correlated with an unusual 3′ acceptor strand hairpin in $tRNA^{Gln}$ [38]. Thus, this hairpin may not form in the TrpRS and TyrRS complexes.

Model building has shown that the active-site crevice in TrpRS is complementary to the undistorted 3′ terminus of $tRNA^{Phe}$. It requires minimal distortion to fit the anticodon close to a site in the helical domain on the other monomer that has been identified, from sequence homology to GlnRS and its high mobility, as a potential anticodon-binding site. It seems likely, therefore, that the anticodon of $tRNA^{Trp}$ binds across the dimer interface from the site of acyl transfer to the acceptor stem, as shown by protein-engineering experiments for TyrRS [43].

The fact that we see essentially a completely ordered TrpRS structure with close structural homologies throughout the three ordered domains of TyrRS raises a puzzling conundrum with respect to tRNA binding. A truncated TyrRS lacking residues 318–417 has essentially the same structure as that illustrated in Fig. 4 [44] and is devoid of tRNA-binding and tRNA-aminoacylation activities [45]. This presents three interesting possibilities: first, the modes of anticodon binding may be rather different in the two enzymes, despite the strong structural similarity between the TrpRS structure and what we can see of the TyrRS structure; second, the contribution of the disordered fragment of TyrRS to tRNA binding may involve wrapping back around the Rossmann fold and into the dimer interface, as we see for the carboxy-terminal helix in TrpRS; and third, the site for anticodon binding suggested by the previous analysis may be incomplete in both enzymes, and the ordered interaction of the carboxy-terminal helix in the dimer interface in this TrpRS structure may undergo a significant rearrangement upon tRNA binding, completing an anticodon-binding site near the helical domain. Before solving the TrpRS·$tRNA^{Trp}$ structure we cannot distinguish between these three possibilities.

Intersubunit communication and acyl transfer

A potential link exists between intersubunit tRNA binding and the conformational isomerism between TrpRS and TyrRS (Figs 4 and 6). Binding of tRNA to sites on two different subunits means that the increase in k_{cat} signaled by correct *versus* mutant anticodon binding observed for $tRNA^{Trp}$ [40] must be communicated across the dimer interface to the active site on the other monomer. The relative orientations of the helical domains differ by ~45° in the two enzymes, which would certainly affect tRNA binding differently, with potential impact on the competence for acyl transfer.

In light of the extensive structural homology and especially the high degree of conservation in the active sites of TrpRS and TyrRS, it would be unusual if the two enzymes did not share other functional aspects also. One such aspect centers on the dramatic difference between the KMSKS loop conformations in the two enzymes. A conformation for this loop similar to that in our TrpRS structure has been demonstrated to occur in TyrRS during the transition-state for amino acid activation, as assayed by pyrophosphate exchange [37]. The comparison in Fig. 6 suggests an answer to the obvious question, do other parts of the TyrRS structure undergo concerted conformational rearrangement during the movement of its KMSKS loop? The configuration of this loop is coupled directly to the configuration of the helical domain in the two structures compared in Figs 4 and 6. Changing the conformation of the helical domain may also destabilize the carboxy-terminal helix interaction across the dimer interface, and thereby loosen it from its moorings, giving rise to the disorder observed in the TyrRS structure. These changes could provide the rudiments of a mechanism for conformational coupling between the site of amino acid activation and the helical domain and its associated site for anticodon binding on the other subunit. There is a clear need to investigate the structures of other conformers of both enzymes to establish the validity and possible details of such a mechanism.

Conformational changes do affect the TrpRS dimer interface during catalysis. Trp91 faces into the dimer interface in our structure, so the fluorescence changes (C Hogue, H Xue and A Szabo, personal communication) and mutation [46] of the corresponding Trp92 in *B. subtilis* TrpRS unequivocally demonstrate changes in the interface upon Trp-5′AMP synthesis.

Because the Trp-5′AMP was synthesized within the preexisting tetragonal crystal lattice after changing the mother liquor, and because we have grown monoclinic crystals under conditions where Trp-5′AMP synthesis is not accompanied by acyl transfer to a second ATP [19], the TrpRS conformation described here must differ somehow from that normally associated with the bound adenylate intermediate. The 18 Å envelope structure of monoclinic TrpRS [47] differs from that reported here in the apparent orientation of the distal helical domains [19]. Moreover, time-resolved fluorescence of 5-hydroxy-Trp-5′AMP in *B. subtilis* TrpRS shows two components, indicating that at least two different conformations of the adenylate complex exist in solution [33]. A consistent possibility is that TrpRS can adopt a conformation closer to that observed for TyrRS, and *vice versa*, and thus the two observed conformations resemble different states assumed by both enzymes during catalysis. Solving other crystal forms of TrpRS, including TrpRS·tRNA complex crystals [48], should clarify the role of conformational changes in intersubunit communication and acyl transfer.

TrpRS and TyrRS are close evolutionary siblings

Class I aaRS sequences have been subdivided into two [2] or more [12,49] subclasses. The chief distinguishing feature is a third 'signature' sequence, WCISR, which is

present in six of the ten class I enzymes and may be correlated with a hydrolytic 'proofreading' function [2]. The four enzymes that lack this sequence are GlnRS, GluRS, TyrRS and TrpRS. Without structure-based alignments, the minimal sequence homology between TrpRS and TyrRS appeared insufficiently compelling to justify grouping them into an additional subclass [13]. There are, however, good structural reasons for considering them separately, as proposed by others [12,49]. TyrRS and TrpRS are dimers whereas GlnRS and GluRS are monomers. Anticodon-binding domains in TrpRS and TyrRS, though incompletely known at present ([43,50–52] and this work), are likely to resemble each other more closely than the dual β-barrels used for anticodon binding by GlnRS.

Surprisingly little similarity exists among any of the amino acid sequence alignments previously proposed for TrpRS and TyrRS [13,49] and that implied by the superpositions shown in Fig. 4. Structure-based alignment [13] is therefore an essential requisite for meaningful sequence comparisons for proteins as distantly related as these. The relatedness of the structurally aligned sequences, of the same length and evaluated without admitting gaps, is highly significant. Overall, of 272 residues superimposed by orienting the three domains as illustrated in Fig. 4, 36 (13%) are identical. The aligned sequences had a Z-score of 9, when compared with 100 randomized sequences of the same length using the BESTFIT routine [53]. This implies a near certainty that the two aligned sequences are related and indicates that the evolutionary divergence of TrpRS and TyrRS from a common ancestor probably occured at a more recent stage of aaRS molecular evolution than that between most other pairs of synthetases, excepting the GlnRS/GluRS and AsnRS/AspRS pairs. Indeed, it is difficult to imagine two enzymes so critically dependent on having different specificities whose structures are so nearly the same.

Biological implications

Translation of the genetic code is arguably the earliest manifestation of biology as we know it. The specificity, mechanisms and evolution of present-day aminoacyl-tRNA synthetases (aaRSs) therefore pose fundamental structural questions. Because they discriminate very efficiently between competing amino acid substrates within a nearly invariant tertiary structural framework, they also provide excellent examples of how enzyme active sites are engineered for specific binding.

We show that *Bacillus stearothermophilus* tryptophanyl-tRNA synthetase (TrpRS) strongly resembles tyrosyl-tRNA synthetase (TyrRS) throughout both structures. Sequence comparisons between enzymes so distantly related are hazardous without structural superposition to identify the gaps. However, the structure-based alignment reveals a highly significant number of identities, providing evidence for descent from a common ancestor of aaRS coding for two rather different amino acids.

The TrpRS active site contains the activated amino acid Trp-5′AMP, and reproduces the specific interactions with adenine, ribose, and α amino moieties previously implicated in catalysis by TyrRS. Unexpectedly, the two enzymes also use nearly identical side-chain arrangements for amino acid recognition, providing an unparalleled comparison of how aaRSs achieve specificity for different amino acids. Specificity for tryptophan *versus* tyrosine involves a subtle tertiary structural switch to reposition a conserved aspartate (132 in TrpRS and 176 in TyrRS). Two prolines immediately preceding Asp132 in the TrpRS specificity-determining helix may effect this switch, as may remodeling to accommodate the six-membered ring of indole.

The TrpRS monomer is too small to bind both ends of $tRNA^{Trp}$. Anticodon $tRNA^{Trp}$ mutants exert a predominant effect on turnover, implicating site–site communication across the dimer interface. Different configurations of very similar modules in TrpRS and TyrRS can be correlated with 'open' (TyrRS) and 'closed' (TrpRS) configurations of the 'KMSKS loop', known to move in the transition state of catalysis by TyrRS, providing a possible link between conformational differences and the communication regulating acyl transfer.

Materials and methods

Crystals and derivatives

Cloning, expression in *E. coli* [54], and growth by microdialysis of tetragonal *B. stearothermophilus* TrpRS crystals with tryptophan (2 mM) and ATP (10 mM) have been described [18,20]. Crystals stabilized for data collection in 3.5 M ammonium sulfate, pH 6.0, were derivatized in the dark at 21°C with 1mM gold chloride for 4 days and 0.1 mM mersalyl acid for 6 h (Au/Hg); and with 30 mM trimethyllead acetate for a week (Pb). Semet (selenomethionyl) TrpRS was made from a methionine auxotroph grown on a defined medium containing selenomethionine [55]. Semet and native TrpRS crystals are isomorphous (space group $P4_32_12$; cell parameters, a=b=60.7 Å, c=233.9 Å with one monomer per asymmetric unit).

Data collection and MIRAS phasing

All derivative and most native datasets were measured from single crystals. Although tetragonal TrpRS crystals diffract to 1.7 Å when stabilized in ammonium sulfate, data were only collected to 2.86 Å because of geometric limitations. Native-1 and Au/Hg data were collected using λ=0.98 Å on the synchroton at Laboratoire pour L'Utilisations des Rayons X (LURE; Orsay, France), integrated, and reduced with MOSFLM [56]. Other datasets were collected on a Rigaku RAXIS IIC imaging-plate system at room temperature with copper radiation and were integrated, scaled and reduced to structure-factor amplitudes using the Rigaku software. The Au/Hg and Pb isomorphous difference Patterson maps were solved by

inspection. Heavy-atom parameters were first refined against centric reflections using REFINE [57] and MLPHARE [58].

Maximum entropy solvent flattening

This procedure [23] involves using a basis set of phased reflections with the best figures of merit, together with a known envelope, to build an exponential model for the electron density of the form:

$$q^{ME}(x)=\exp(\text{basis set } \Sigma\zeta_h e^{2\pi ih\cdot x})$$

consistent with the phase and envelope constraints. Fitting the parameters of this model, ζ_h, leads to extrapolated values for the amplitudes and phases outside the basis set. Because in the absence of an envelope there are as many parameters as constraining data the constraints can be fitted exactly; however, doing so fits noise in the data as well as errors in the constraints. Fitting is therefore stopped when the extrapolated amplitudes outside the basis set best match the observed amplitudes, as indicated by a criterion called the 'log likelihood gain' [24].

Phase and envelope permutation

As phase determination progressed with the non-isomorphous heavy-atom phases, 28 intense reflections outside the basis set possessing weak maximum–entropy extrapolation were phased directly. Successive full and incomplete factorial designs [18] were used to increase the basis set with different phase combinations for three centric and four acentric (tested at 45°, 135°, 225° and 315°) reflections at a time. Similar permutation was carried out for five different binary choices regarding calculation and description of the molecular envelope. Permutation experiments were scored using the log-likelihood gain, and the average differences for experiments evaluated with the same phase choice for each reflection were analyzed by multiple regression least-squares. Student *t*-tests gave significant indications for phases and all six hypotheses regarding the envelope. The resulting phase improvement made it possible to assign positions (hitherto unobtainable) for nine of the ten Se atoms in an isomorphous difference Fourier map for Semet TrpRS crystals [25], thereby incorporating these isomorphous differences into the final experimental phase determination.

Model building and refinement

The polypeptide chain was fitted to the electron density using O [59]. The model, now complete except for residues Arg327 and Arg328, was refined without water molecules using X-PLOR 2.1 and 3.0 [60]. The crystallographic R-factor is 19.7% for all 9631 reflections from 7 Å to 2.86 Å resolution, with an rms deviation from ideal geometry of 0.011 Å for bond lengths and 2.8° for bond angles (Table 1). Individual isotropic B-factors were refined at the end of the refinement. The model stereochemistry was verified in detail using PROCHECK [61]. All categories are 'better' than average; in particular, the residue with the worst stereochemistry is inside the generously allowed region of the Ramachandran plot.

Graphics

Orientations and items to be displayed in all molecular illustrations were initially chosen using MAGE [62]. Figs 2a and 2c were drawn using RIBBONS [63] and processed digitally with Adobe PHOTOSHOP [64]. Fig. 4 was prepared using MOLSCRIPT [65], TRANSVERTER PRO [66] and MACDRAW PRO [67]. Figs 3a, 3b and 9 were prepared using MAXIMAGE [68] and MACDRAW PRO [67].

Atomic coordinates for all atoms have been submitted to the Brookhaven Protein Data Bank and are available from the authors (carter@med.unc.edu).

Acknowledgements: We are grateful to LURE, especially R Fourme and to W Royer for beam time and assistance with data collection. M Rould provided continuing support and valuable discussions. We thank J Rogers for pointing out the sequence homology between TrpRS and the anticodon-binding region in GlnRS, D Söll for numerous discussions, H Xue, A Szabo and C Hogue for communicating results before publication, and R Doolittle for advice on comparing sequences. We also thank C Bruton and T Atkinson for early support of the project. This work was supported by grants from the NIH, NSF and American Cancer Society.

References

1. Schimmel, P. (1987). Aminoacyl-tRNA synthetases: general scheme of structure–function relationships in the polypeptides and recognition of transfer RNAs. *Annu. Rev. Biochem.* **56**, 125–158.
2. Carter, C.W., Jr. (1993). Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases. *Annu. Rev. Biochem.* **62**, 715–748.
3. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347**, 203–206.
4. Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N. & Leberman, R. (1990). A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature* **347**, 249–255.
5. Brick, P., Bhat, T.N. & Blow, D.M. (1989). Structure of tyrosyl-tRNA synthetase refined at 2.3 Å resolution: interaction of the enzyme with the tyrosyl adenylate intermediate. *J. Mol. Biol.* **208**, 83–98.
6. Brunie, S., Zelwer, C. & Risler, J.L. (1990). Crystallographic study at 2.5 Å resolution of the interaction of methionyl-tRNA synthetase from *Escherichia coli* with ATP. *J. Mol. Biol.* **216**, 411–424.
7. Rould, M.A., Perona, J.J., Söll, D. & Steitz, T.A. (1989). Structure of the *E. coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution. *Science* **246**, 1135–1142.
8. Rossmann, M.G., Moras, D. & Olsen, K.W. (1974). Chemical and biological evolution of a nucleotide binding domain. *Nature* **250**, 194–199.
9. Webster, T.A., Lathrop, R.H. & Smith, T.F. (1987). Prediction of a common structural domain in aminoacyl-tRNA synthetases through use of a new pattern-directed inference system. *Biochemistry* **26**, 6950–6957.
10. Hountondji, C., Dessen, P. & Blanquet, S. (1986). Sequence similarities among the family of aminoacyl-tRNA synthetases. *Biochimie* **68**, 1071–1078.
11. Perona, J.J., *et al.*, & Brunie, S. (1991). Structural similarities in glutaminyl- and methionyl-tRNA synthetases suggest a common overall orientation of tRNA binding. *Proc. Natl. Acad. Sci. USA* **88**, 2903–2907.
12. Delarue, M. & Moras, D. (1992). Aminoacyl-tRNA synthetases: partition into two classes. In *Nucleic Acids and Molecular Biology*. **6**, (Eckstein, F. & Lilley, D.M.J., eds), pp. 203–224, Springer-Verlag, Berlin, Heidelberg.
13. Landès, C., *et al.*, & Risler, J.-L. (1995). A structure-based multiple sequence alignment of all Class I aminoacyl-tRNA synthetases. *Biochimie*, in press.
14. Nagel, G.M. & Doolittle, R.F. (1995). Phylogeny of the aminoacyl-tRNA synthetases. *J. Mol. Evol.*, in press.
15. Fersht, A.R. (1987). Dissection of the structure and activity of the tyrosyl-tRNA synthetase by site-directed mutagenesis. *Biochemistry* **26**, 8031–8037.
16. Schimmel, P. & Söll, D. (1979). Aminoacyl-tRNA synthetases: general features and recognition of transfer RNAs. *Annu. Rev. Biochem.* **48**, 601–648.
17. Rould, M.A. & Steitz, T.A. (1992). Structure of the glutaminyl-tRNA synthetase–$tRNA^{Gln}$ complex. In *Nucleic Acids and Molecular Biology*. **6**, (Eckstein, F. & Lilley, D.M.J., eds), pp. 225–246, Springer-Verlag, Berlin, Heidelberg.
18. Carter, C.W., Jr. & Carter, C.W. (1979). Protein crystallization using incomplete factorial experiments. *J. Biol. Chem.* **254**, 12219–12223.
19. Carter, C.W., Jr., Doublié, S. & Coleman, D.E. (1994). Quantitative analysis of crystal growth: tryptophanyl-tRNA synthetase polymorphism and its relationship to catalysis. *J. Mol. Biol.* **238**, 346–365.

20. Coleman, D.E. & Carter, C.W., Jr. (1984). Crystals of *Bacillus stearothermophilus* tryptophanyl-tRNA synthetase containing enzymatically formed acyl transfer product tryptophanyl-ATP, an active site marker for the 3' CCA terminus of tryptophanyl-tRNA^{Trp}. *Biochemistry* **23**, 381–385.
21. Bricogne, G. (1988). A bayesian statistical theory of the phase problem. I. A multichannel maximum-entropy formalism for constructing generalized joint probability distributions of structure factors. *Acta Crystallogr. A* **44**, 517–545.
22. Bricogne, G. (1993). Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives. *Acta Crystallogr. D* **49**, 37–60.
23. Xiang, S., Carter, C.W., Jr, Bricogne, G. & Gilmore, C.J. (1993). Entropy maximization constrained by solvent flatness: a new method for macromolecular phase extension and map improvement. *Acta Crystallogr. D* **49**, 193–212.
24. Bricogne, G. & Gilmore, C.J. (1990). A multisolution method of phase determination by combined maximization of entropy and likelihood I theory, algorithms and strategy. *Acta Crystallogr. A* **46**, 248–297.
25. Doublié, S., Xiang, S., Gilmore, C.J., Bricogne, G. & Carter, C.W., Jr. (1994). Overcoming non-isomorphism by phase permutation and likelihood scoring: solution of the TrpRS crystal structure. *Acta Crystallogr. A* **50**, 164–182.
26. Lapthorne, A.J., *et al.*, & Isaacs, N. (1994). Crystal structure of human chorionic gonadotropin. *Nature* **369**, 455–461.
27. Barstow, D.A., Sharman, A.F., Atkinson, T. & Minton, N.P. (1986). Cloning and complete nucleotide sequence of the *Bacillus stearothermophilus* tryptophanyl-tRNA synthetase gene. *Gene* **46**, 37–42.
28. Hall, C.V., vanCleemput, M., Muench, K.H. & Yanofsky, C. (1982). The nucleotide sequence of the structural gene for *Escherichia coli* tryptophanyl-tRNA synthetase. *J. Biol. Chem.* **257**, 6132–6136.
29. Edelhoch, H. & Osborne, J.C., Jr. (1976). The thermodynamic basis of stability of proteins, nucleic acids, and membranes. In *Advances in Protein Chemistry.* **30**, (Anfinsen, C.B., Edsall, J.T. & Richards, F.M., eds), pp. 183–250, Academic Press, New York.
30. Rossmann, M. & Argos, P. (1976). Exploring structural homology of proteins. *J. Mol. Biol.* **105**, 75–95.
31. Joseph, D.R. & Muench, K. (1971). Tryptophanyl-transfer ribonucleic acid synthetase of *Escherichia coli.* 1. Purification of the enzyme and of tryptophan transfer ribonucleic acid. *J. Biol. Chem.* **246**, 7602–7609.
32. Burley, S.K. & Petsko, G.A. (1985). Aromatic–aromatic interaction: a mechanism of protein structure stabilization. *Science* **229**, 23–28.
33. Hogue, C.W.V. & Szabo, A.G. (1993). Characterization of aminoacyl-adenylates in *B. subtilis* tryptophanyl-tRNA synthetase by the fluorescence of tryptophan analogs 5-hydroxytryptophan and 7azatryptophan. *Biophys. Chem.* **48**, 159–169.
34. Fishel, L.A., *et al.*, & Scholes, C.P. (1991). Compound I radical in site-directed mutants of cytochrome c peroxidase as probed by electron paramagnetic resonance and electron-nuclear double resonance. *Biochemistry* **30**, 1986–1996.
35. Kim, H.Y., Ghosh, G., Schulmann, L.H., Brunie, S. & Jakubowski, H. (1993). The relationship between synthetic and editing functions of the active site of an aminoacyl-tRNA synthetase. *Proc. Natl. Acad. Sci. USA* **90**, 11553–11557.
36. Garret, M., *et al.*, & Bonnet, J. (1991). A mammalian tryptophanyl-tRNA synthetase shows little homology to prokaryotic synthetases but near identity with mammalian peptide chain release factor. *Biochemistry* **30**, 7809–7817.
37. Fersht, A.R., Knill-Jones, J.W., Bedouelle, H. & Winter, G. (1988). Reconstruction by site-directed mutagenesis of the transition state for the activation of tyrosine by the tyrosyl-tRNA synthetase: a mobile loop envelopes the transition state in an induced-fit mechanism. *Biochemistry* **27**, 1581–1587.
38. Perona, J., Rould, M.A. & Steitz, T.A. (1993). Structural basis for transfer RNA aminoacylation by *Escherichia coli* glutaminyl-tRNA synthetase. *Biochemistry* **32**, 8758–8771.
39. Fersht, A.R., *et al.*, & Winter, G. (1985). Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314**, 235–238.
40. Rogers, M.J., Adachi, T., Inokuchi, H. & Söll, D. (1992). Switching tRNA^{Gln} identity from glutamine to tryptophan. *Proc. Natl. Acad. Sci. USA* **89**, 3463–3467.
41. Himeno, H., Hasegawa, T., Asahara, H., Tamura, K. & Shimizu, M. (1991). Identity determinants of *E. coli* tryptophan tRNA. *Nucleic Acids Res.* **19**, 6379–6382.
42. Xue, H., Shen, W., Giegé, R. & Wong, J.T.-F. (1993). Identity of tRNA^{Trp}: identification and evolutionary conservation. *J. Biol. Chem.* **268**, 9316–9322.
43. Bedouelle, H. & Winter, G. (1986). A model of synthetase/transfer RNA interaction as deduced by protein engineering. *Nature* **230**, 371–373.
44. Brick, P. & Blow, D.M. (1987). Crystal structure of a deletion mutant of a tyrosyl-tRNA synthetase complexed with tyrosine. *J. Mol. Biol.* **194**, 287–297.
45. Waye, M.M.Y., Winter, G., Wilkinson, A.J. & Fersht, A.R. (1983). Deletion mutagenesis using an 'M13 splint': the N-terminal structural domain of tyrosyl-tRNA synthetase (*B. stearothermophilus*) catalyses the formation of tyrosyl adenylate. *EMBO J.* **2**, 1827–1829.
46. Chow, K.-C., Xue, H., Shi, W. & Wong, J.T.-F. (1992). Mutational identification of an essential tryptophan in tryptophanyl-tRNA synthetase of *Bacillus subtilis. J. Biol. Chem.* **267**, 9146–9149.
47. Carter, C.W., Jr., Crumley, K.V., Coleman, D.E., Hage, F. & Bricogne, G. (1990). Direct phase determination for the molecular envelope of tryptophanyl-tRNA synthetase from *Bacillus stearothermophilus. Acta Crystallogr. A* **46**, 57–68.
48. Carter, C.W., Jr, Green, D.C., Toomim, C. & Betts, L. (1985). Two-step purification of tryptophan-accepting tRNA from *Bacillus stearothermophilus. Anal. Biochem.* **151**, 515–519.
49. Nagel, G.M. & Doolittle, R.F. (1991). Evolution and relatedness in two aminoacyl-tRNA synthetase families. *Proc. Natl. Acad. Sci. USA* **88**, 8121–8125.
50. Labouze, E. & Bedouelle, H. (1989). Structural and kinetic bases for the recognition of tRNA^{Tyr} by tyrosyl-tRNA synthetase. *J. Mol. Biol.* **205**, 729–735.
51. Bedouelle, H. (1990). Recognition of tRNA^{Tyr} by tyrosyl-tRNA synthetase. *Biochimie* **72**, 589–598.
52. Bedouelle, H., Guez-Ivanier, V. & Nageotte, R. (1993). Discrimination between transfer-RNAs by tyrosyl-tRNA synthetase. *Biochimie* **75**, 1099–1108.
53. Genetics Computer Group (1991). Program Manual for the GCG Package. University of Wisconsin, Madison, WI.
54. Carter, C.W., Jr (1988). Cloning heterologous genes into *E. coli* for enzyme production and crystal growth: problems of expression and microheterogeneity. *J. Crystal Growth* **90**, 168–179.
55. Doublié, S. & Carter, C.W., Jr (1992). Preparation of selenomethionyl protein crystals. In *Crystallization of Proteins and Nucleic Acids: A Practical Approach.* (Ducruix, A. & Giegé, R., eds), pp. 311–317, Oxford University Press, Oxford.
56. Leslie, A. (1989). MOSFLM. Joint CCP4 and ESF-EACBM newsletter 26 on protein crystallography.
57. CCP4 (1991). The SERC(UK) Collaborative computing project No 4: A suite of programs for protein crystallography. Daresbury Laboratory, Warrington, UK.
58. Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In *Isomorphous Replacement and Anomalous Scattering* (Wolf, W., Evans, P.R. & Leslie, A.G.W., eds), pp. 80–86, SERC Daresbury Laboratory, Warrington, UK.
59. Jones, T.A., Bergdoll, M. & Kjeldgaard, M. (1990). O: a macromolecule modeling environment. In *Crystallographic and Modelling Methods in Molecular Design.* (Ealick, S. & Bugg, C., eds), pp. 189–199, Springer-Verlag, New York.
60. Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R-factor refinement by molecular dynamics. *Science* **235**, 458–460.
61. Morris, A.L., MacArthur, M.W., Hutchinson, E.G. & Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. *Proteins* **12**, 345–364.
62. Richardson, D.C. & Richardson, J.S. (1992). Kinemage, a tool for scientific communication. *Protein Sci.* **1**, 3–9.
63. Carson, M. (1992). *Ribbons, Version 2.0.* University of Alabama at Birmingham.
64. Adobe Systems (1994). Photoshop. Adobe Systems Inc., Mountain View, CA.
65. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
66. Techpool Software, I (1993). Transverter Pro. Techpool Software, Inc., Cleveland, OH.
67. Claris, C (1992). MacDraw Pro. Claris Corporations, Santa Clara, CA.
68. Rould, M. (1990). MAXIMAGE: a program for rapidly generating photo-ready FRODO and SZAZAM images and black-and-white figures. Yale University, New Haven, CT.

Received: **26 Sep 1994**; revisions requested: **13 Oct 1994**; revisions received: **26 Oct 1994**. Accepted: **8 Nov 1994**.